EXHIBIT F

UNITED STATES DISTRICT COURT
CENTRAL DISTRICT OF CALIFORNIA
WESTERN DIVISION

SANDOZ INC., Plaintiff, vs. AMGEN INC., Defendant.	Case No. 2:22-cv-05326

Expert Report of

Robert D. Gibbons

Blum-Riese Professor of Biostatistics
University of Chicago

May 10, 2023

address limitations associated with RCTs. RCTs may exhibit a higher degree of patient compliance with treatments under investigation, relative to the level of compliance seen in real-world use. Patient compliance is an especially important consideration when comparing OBI and PFS. For this reason, observational studies based on real-world data that reflect realistic levels of patient compliance can offer meaningful insights in this context.

12. The three studies mentioned in the Complaint (both Amgen studies and the 2021 Sandoz Retrospective Study) individually and collectively provide compelling statistical evidence of a lower incidence of FN associated with the use of OBI. While both Amgen studies reveal a statistically significantly lower incidence of FN associated with the use of OBI, the 2021 Sandoz Retrospective study finds an effect of similar magnitude that is not statistically significant. Based on my review, I conclude that the 2021 Sandoz Retrospective study was underpowered to perform a suitable comparison of FN rates associated with the use of OBI and PFS, even though the magnitude of the estimated effect reported in the study is consistent with the magnitude of the statistically significant effects reported in the Amgen studies. Further, by synthesizing the data of the three studies in a meta-analysis, I find that collectively the studies reveal a statistically significant 35% increase in the incidence of FN associated with the use of PFS and other comparators relative to OBI. My evaluation of the studies individually and collectively allows me to conclude that the claims in Amgen's marketing materials regarding OBI are consistent with the results of the Amgen studies. The Amgen studies provided valuable statistical insights of scientific interest, such that it was reasonable for Amgen to rely on the results of these studies in issuing the challenged marketing materials.

13. I have reviewed the specific concerns regarding the Amgen studies raised by Sandoz in their Complaint,[3] and find that none of these concerns undermine the statistical evidence of an association between the use of OBI and a lower incidence of FN. In fact, some of these concerns are already addressed by Amgen's own studies, as well as the Sandoz study.[4] For example, criticisms regarding the failure of the 2019 Amgen Retrospective Study to control for potential observed confounders are addressed by both the 2021 Amgen Prospective Study and the 2021 Sandoz Retrospective Study. Both of these studies calculated the reduction in the FN incidence rates associated with OBI after controlling for observed confounders.

---

[3] See Complaint, *Sandoz Inc. v. Amgen Inc. and Amgen Manufacturing Limited*, August 1, 2022 ("Complaint").
[4] Ali McBride et al., "Economic and Clinical Outcomes of Pegfilgrastim via Prefilled Syringe vs On-Body Injector: A Real-World Data Analysis," *Journal of Managed Care and Specialty Pharmacy*, 27(9), September 2021, pp. 1230–1238 ("2021 Sandoz Retrospective Study").

These adjusted analyses show *even larger* reductions in the FN incidence rate relative to the 2019 Amgen Retrospective Study which did not perform such adjustments. This provides even further statistical evidence in support of a reduction in the FN incidence rate associated with OBI.

IV. Relevant Scientific Background

A. Explanation of Relevant Methods Used in Statistical Analysis

14. Statistics is a field of science concerned with the testing of hypotheses, estimation of summary measures such as the mean value (e.g., the average value in the case of a normal distribution) or the strength of association between two or more variables, and the characterization of uncertainty.

15. For example, we may be interested in understanding the efficacy of a new medical treatment for FN. To this end, we might conduct a study that compares FN outcomes of patients receiving this new medical treatment (treatment group) to those receiving alternate treatments or no treatment (control group). At the end of the study, we are interested in estimating the rate of developing FN in patients treated with the new medication compared to those who are not. This is an example of statistical estimation.

16. Since these observed rates are estimates of the population rates, we must also characterize our uncertainty in the true population values. To do this we can compute the standard error of the estimate, which is a statistical estimate of the uncertainty, and/or the confidence interval. The confidence interval allows us to construct an upper and lower bound for plausible values of the statistic of interest. The interpretation of the confidence interval is that had we repeated the experiment multiple times, 95% of the time the true population rate would lie within the computed confidence interval. This does not mean that the true population value lies within the confidence interval for any single experiment with the given level of confidence. As the sample size (i.e., number of subjects) increases, all other things being equal, the confidence interval decreases in width. As the size of the sample from the population goes to infinity, the confidence interval has no width and is simply the population rate or the point estimate of the statistic of interest.

17. To this point we have estimated the rate of FN at the end of the study for the two groups (treated and control patients) and have characterized our uncertainty in these two

have stringent entrance criteria and treatment protocols that may have very little to do with the type of patients that will ultimately receive the treatment and the dosage and duration that is indicative of real-world medical practice. Observational studies are often more representative of real-world medical practice, can be much larger than RCTs, and may therefore be more generalizable to the population of interest. Smoking and lung cancer is a good example where large-scale observational studies ultimately led to the inference that smoking cigarettes causes lung cancer.[10] For these reasons, both RCTs and observational studies are scientifically important, particularly in medical research.

C. Common Statistical Terms Used in This Report

29. Related statistical terms that will be discussed in further detail in the report are defined here:

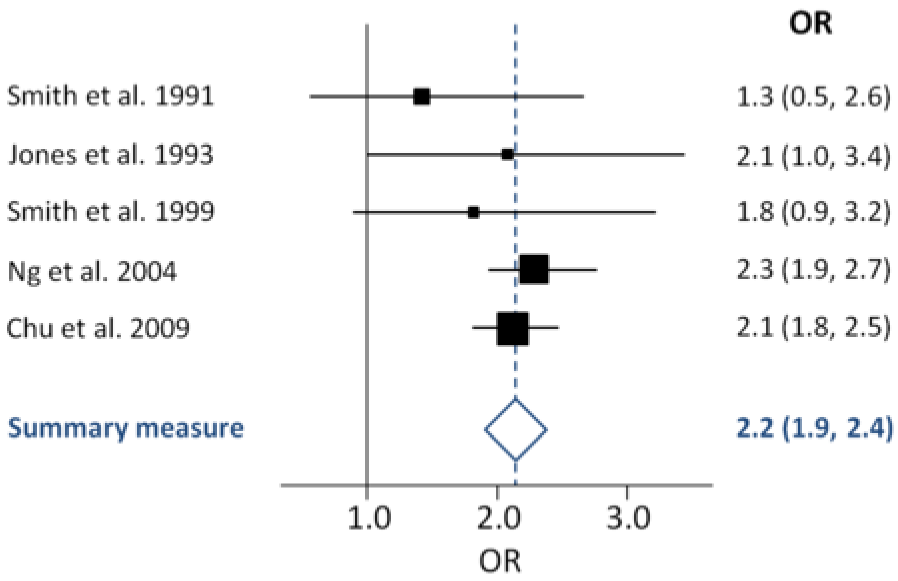
- **Sample Variance** – A measure of spread of the distribution. It is computed as the average squared deviation from the mean, using n-1 rather than n (the sample size) in the denominator.

- **Sample Standard Deviation** – A measure of spread of the distribution, computed as the square root of the sample variance. The larger the standard deviation, the larger the deviations of the individual values from the average value (mean).

- **Relative Risk (RR)** – The ratio of the probability or incidence of an event occurring in the treated (exposed) group to the probability or incidence of an event occurring in the control group. For example, if the rate of death in an exposed group is 20% and the rate of death in the control group is 10%, then the relative risk is RR = (0.2)/(0.1) = 2.

- **Forest Plot** – A graphical presentation of a meta-analysis in which the individual study effect size (standardized mean difference, RR, or other risk measures such as odds ratio (OR) or hazard rate (HR)) is displayed for each study along with its 95% confidence interval, and a pooled estimate of the overall effect size and its 95% confidence interval are displayed at the bottom of the plot. The following is an example of a Forest Plot: (a) the dark squares are the study-specific ORs with the size of the square proportional to the weights used in the meta-analysis (typically a function of uncertainty which is proportional to sample size), (b) the 95% confidence limits for the study ORs are described as the lines intersecting the squares, (c) a summary measure is displayed as a diamond at the bottom, with the 95% confidence interval displayed as the lateral tips of the diamond, and (d) a solid vertical line denotes OR=1.0 (no effect).

[10] U.S. Department of Health and Human Services, "The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General," U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2014, p. 21.



- **Propensity Score Matching** – A statistical matching strategy in which a series of potential confounding variables are matched in probability between those who did and did not receive the intervention, treatment, or exposure of interest.

- **Marginal Structural Log-Binomial Regression Model** – For cross-sectional data, the marginal structural log-binomial model is a form of propensity score matching that uses inverse probability weighting to balance exposure arms (e.g., treated and control). Marginal structural models can also be used to adjust for dynamic confounding in longitudinal data with a time-varying exposure or treatment variable. The marginal structural model replaces covariate adjustment of the outcome with a propensity score model for the exposure or treatment. Instead of matching used in propensity score algorithms, which can lead to loss of informative parts of the sample, the marginal structural model uses the propensity score as a weight and retains all participants in the analysis. As such, it can be less biased than propensity score matching when the underlying confounder distributions may not have common support.

- **Sensitivity Analysis** – Studies are typically defined with a prespecified primary endpoint and statistical analysis plan. In some cases, we conduct a sensitivity analysis to determine how sensitive the results are to changes in statistical specifications. For example, we might conduct a sensitivity analysis to compare under the assumption that the underlying data observes a specific statistical distribution as opposed to another. In observational studies, we might conduct a sensitivity analysis to determine if the results change if, for example, we restrict the sample to participants with a particular form of cancer (e.g., breast cancer).

D. Types of Studies Used in Medical Research

30. In this section, I discuss certain types of studies commonly used in medical research to analyze the relationship between a treatment of interest and its outcome.

1. RCTs

31. RCTs are commonly used to isolate the causal effect of a treatment under investigation. RCTs allow the researcher to establish causality in the most scientifically rigorous way by minimizing bias induced by confounding variables. That is, an RCT will produce unbiased results even in the presence of confounding variables, due to the randomization of participants to the treatment or control group. A double-blind RCT, in which the patients and raters are blind to the treatment received, further reduces potential bias. In the context of medical science, clinical trials are the highest standard required to establish a causal link between a treatment of interest and the outcome. For this reason, they are an important part of the FDA's process of approving new drugs.

32. However, as noted earlier, RCTs have certain limitations that may complicate generalizability of results to the broader population. The study protocol in an RCT might exclude subjects who would receive a drug in a real-world setting, leading to concerns about sample representativeness. Further, subject compliance to the treatment in an RCT may be higher relative to the level of compliance that would be observed in actual clinical practice. These limitations, among others, create the need for other kinds of studies.

2. Observational Cohort Studies

33. As mentioned above, observational studies are typically much larger than RCTs and more representative of actual medical practice, which may allow for greater generalizability of results in certain applications. In pharmacoepidemiology, observational cohort studies define a population of interest (e.g., factory workers) and examine the relationship between an exposure (e.g., asbestos) and an outcome (e.g., cancer).

34. Cohort studies can be retrospective or prospective. Retrospective cohort studies rely on data that have already been collected for purposes other than the study itself in order to study associations between variables of interest. For example, if the data used in a retrospective study contain clinical information but not demographic information, the retrospective study will not be able to analyze the effect of age on the outcome variable of interest, while it may still be able to analyze the effect of certain comorbidities or medical conditions on this outcome variable.

35. Prospective cohort studies generally have greater validity than retrospective cohort studies. In prospective cohort studies, a researcher has more control over the study design,

including protocols for data collection, information recording, desired sample size, studied endpoints, and data analyses to be conducted. Depending on the goals of the prospective study, a researcher can adapt the study design so that the limitations of the cohort study are minimized.

36. Cohort studies, both retrospective and prospective, can have limitations, such as bias associated with selection effects of patients to treatments and variation in the quality of available endpoints (e.g., outcome measures of interest related to FN). Confounding can be between individuals or even within individuals where changes in the severity of the underlying disease increases the likelihood of treatment and the adverse event of interest, providing the appearance of an association.[11] In general, as described above, potential confounding factors regarding the exposure-outcome relationship are controlled by *matching* and/or *statistical adjustment*. It is important to note, however, that matching and adjustment are based on *observed* confounders only, and residual confounding may remain based on unmeasured or unobserved factors.

3. Meta-Analyses

37. Meta-analysis is a statistical method that synthesizes the results from a number of studies, increasing sample size and statistical power of the resulting analysis. There are random-effect and fixed-effect meta-analysis models.[12] Fixed-effect models assume that there is a common treatment effect across all studies, whereas random-effect models allow for heterogeneity in treatment effect across studies. Random-effect models are designed to allow the comparison of studies that may differ in terms of primary endpoints, exclusion criteria, study design, comparator groups, or any other minor discrepancy that the protocols may have. Such differences are absorbed by the treatment heterogeneity component of the model. More recently, mixed-effects regression models have been developed for meta-analysis and have many advantages over the classical approaches. However, any form of meta-analysis is an observational study of studies. Differences between study designs and endpoints can bias results and corresponding p-values, point estimates, and confidence intervals. Policymakers routinely rely on meta-analyses to inform public policies. For

[11] For example, the relationship between antidepressants and suicide is confounded with depression, which increases the likelihood of both treatment with antidepressants and suicide risk, which will produce a non-causal association between antidepressants and suicide. See Robert D. Gibbons and Anup K. Amatya, *Statistical Methods for Drug Safety*, 1st Edition, (Chapman and Hall/CRC Press, 2015), at p. 4.

[12] The classical fixed-effect model is the Mantel-Haenszel method, and the classical random-effect model is the DerSimonian and Laird model.

example, in 2004, the FDA issued a "black box warning" on the link between antidepressants and suicide in children based on a meta-analysis of 24 RCTs and extended it to young adults under age 25 based on a meta-analysis of 372 RCTs.[13]

V. Observational Studies Can Provide Valuable Insights into Real-World Treatment Performance, Especially in Comparing OBI and PFS

38. Previously, I discussed the advantages and limitations of different types of studies that can serve as evidence in the field of pharmacoepidemiology. As indicated above, while RCTs may be well-suited to evaluate the causal effect of a given treatment, RCTs may have limitations in demonstrating performance of a treatment in a real-world setting.

39. A real-world setting contains various features that are not present in the highly controlled and moderated environment of an RCT.[14] One important limitation of RCTs is the fact that patient compliance with a treatment regimen being studied can be very high relative to the compliance levels in a typical use case of a product.[15] This may happen because subjects usually try to follow the guidelines of the studies in which they knowingly enroll. As a consequence of this elevated compliance, the results may not provide an accurate characterization of the performance of this treatment in a real-world setting where compliance may not be as high.[16] For example, a treatment protocol may be difficult or cumbersome for patients to follow on their own, and hence even if an RCT were to establish the effectiveness of the treatment, it may not be able to capture the effectiveness of the treatment in a real-world setting where patients are disinclined to adopt it.[17]

40. RWE generated through observational studies can offer useful insights into the real-world performance of the treatment under investigation. While RCTs essentially guarantee

---

[13] "Suicidality in Children and Adolescents Being Treated With Antidepressant Medications," FDA, available at https://www.fda.gov/drugs/postmarket-drug-safety-information-patients-and-providers/suicidality-children-and-adolescents-being-treated-antidepressant-medications; Richard A. Friedman and Andrew C. Leon, "Expanding the Black Box — Depression, Antidepressants, and the Risk of Suicide," *The New England Journal of Medicine*, 356, 2007, pp. 2343–2346.
[14] David C. Klonoff, "The Expanding Role of Real-World Evidence Trials in Health Care Decision Making," *Journal of Diabetes Science and Technology*, 14(1), 2020, pp. 174–179 ("Klonoff (2020)"), at p. 175.
[15] Klonoff (2020), p. 176 ("Adherence can be high in a clinical trial that measures efficacy…whereas in the real world adherence may be low and outcomes may be less favorable.").
[16] Melissa Roberts and Gary T. Ferguson, "Real-World Evidence: Bridging Gaps in Evidence to Guide Payer Decisions," *PharmacoEconomics – Open*, 5(1), 2021, pp. 3–11 ("Roberts and Ferguson (2021)"), p. 5 ("Patients in general, and their medication-taking behavior specifically, are closely monitored in RCTs; thus, adherence, persistence, and compliance outcomes may not reflect actual, real-world patient behavior.").
[17] See Faye M. Pais et al., "Influence of Clinical Factors and Exclusion Criteria on Mortality in ARDS Observational Studies and Randomized Controlled Trials," *Respiratory Care*, 63(8), 2018, pp. 1060–1068.

which tend to be underrepresented in clinical trials.[29] Thus, a randomized trial missing these underrepresented participants could lead to a biased estimate of the difference between FN rates in OBI and PFS cohorts—i.e., a smaller difference than is actually seen in the real world due to unrepresentative treatment compliance in the PFS arm.

44. Thus, to the extent that compliance with the prescribed administration is significantly lower when using PFS relative to OBI, observational studies can provide meaningful insight into the real-world performance of the treatment. Healthcare providers concerned about the well-being of their patients would be especially interested in results of such observational studies, especially as it concerns prescribing decisions.

VI. The Amgen Studies and the Sandoz Study, Both Individually and Collectively, Provide Statistical Evidence of a Lower Rate of FN Associated with the Use of OBI

45. In this section, I review and discuss the results of the three observational studies mentioned in the Complaint that compare FN rates associated with the use of OBI versus PFS, or the use of OBI versus other prophylaxis options (including PFS). The 2019 Amgen Retrospective Study is a retrospective cohort study that used MarketScan medical claims data to compare rates of FN between OBI and PFS administration of Neulasta. The 2021 Sandoz Retrospective Study reported results based on a similar study that used propensity score matching in an attempt to balance the PFS and OBI groups in terms of measured potential confounders. The 2021 Amgen Prospective Study is a prospective cohort study that compared OBI and other prophylaxis methods, including pegfilgrastim PFS, in terms of a prospective biologically objective FN endpoint. In addition to calculating the raw FN incidence rates, the 2021 Amgen Prospective Study performed analyses which adjusted the FN incidence rates for potential confounders using a marginal structural model.

46. I explain the results of each of these studies individually in detail below and the extent to which they provide statistical evidence of a lower incidence of FN associated with OBI. I explain how the claims in the Amgen marketing materials regarding the performance of OBI relative to PFS, or the performance of OBI relative to other prophylaxis options (including PFS), are consistent with the results of the Amgen studies. Finally, I conduct a meta-analysis

---

[29] Manuel A. Ma et al., "Minority Representation in Clinical Trials in the United States: Trends Over the Past 25 Years," *Mayo Clinic Proceedings*, 96(1), 2021, pp. 264–266; Ryan W. Huey et al., "Patient-Reported Out-of-Pocket Costs and Financial Toxicity During Early-Phase Oncology Clinical Trials," *Oncologist*, 26(7), 2021, pp. 588–596.

of these studies, which finds that they collectively provide compelling statistical evidence of a lower incidence of FN associated with OBI.

A. The 2019 Amgen Retrospective Study

47. The 2019 Amgen Retrospective Study analyzed a total of 35,856 Neulasta cycles (9,395 for OBI and 26,461 for PFS).[30] There were 126 FN cases for OBI (1.3%) and 455 FN cases for PFS (1.7%).[31] Data were derived from the MarketScan medical claims database. The study period was 1/1/16 – 9/30/18.[32] The patient selection period was 01/01/2017 – 05/31/2018 and the patients were followed for 1 to 12 months following the start of their first chemotherapy cycle.[33]

48. In the 2019 Amgen Retrospective Study, Amgen's primary analysis found a statistically significant (p = 0.01) 31% FN rate increase in patients treated with PFS relative to OBI ((1.7–1.3)/1.3 = 0.31).[34] I have computed the relative risk (RR) and associated 95% confidence interval based on the reported data and find RR = 0.78 (95% confidence interval (CI) = 0.64, 0.95; p = 0.01 (2-tailed)).[35]

49. In addition to the primary analysis, the 2019 Amgen Retrospective Study also reported the percentage of times that treatment was initiated on the day following the end of chemotherapy as recommended.[36] For OBI, the percentage was 98.5% (CI = 98.2%, 98.7%) and for PFS it was 58.4% (CI = 57.8%, 59.0%), indicating that compliance with the prescribed administration was 69% greater for OBI than PFS.[37] This is important because it suggests that compliance with the treatment is a likely mechanism for the significant reduction in FN associated with the use of OBI versus PFS.

---

[30] 36,421 cycles satisfied the protocol inclusion criteria and for 35,856 of these cycles, claims data allowed the researchers identify the presence or absence of FN. See Defendant's Supplemental Responses and Objections to Plaintiff's First Set of Interrogatories to Defendant Amgen Inc. (Nos. 1–10), *Sandoz Inc. v. Amgen Inc. and Amgen Manufacturing Limited*, March 7, 2023 ("Amgen ROG"), p. 29; AMG10939-0000304080, tab "5a_FN&Cost_Cycle."

[31] Amgen ROG, p. 53.

[32] Amgen ROG, p. 26.

[33] Amgen ROG, p. 27.

[34] AMG10939-0000303957–84 at 72.

[35] I use Open Epi's interface to produce specific calculation in this report. I provide details on inputs and outputs that I use in these calculations as part of my backup materials. Dean AG, Sullivan KM, Soe MM. OpenEpi: Open Source Epidemiologic Statistics for Public Health, available at www.OpenEpi.com. See Workpaper 1, Row 001, and my backup materials. The percentage increase in the FN incidence rate for the PFS patients relative to the OBI patients is calculated as 100*(1/RR–1). The percentage increase reported in the marketing materials (31%) differs slightly from the percentage increase calculated using this formula (28%), but I have confirmed that this difference is simply due to rounding error.

[36] AMG10939-0000303957–84 at 71.

[37] AMG10939-0000303957–84 at 71. See my backup materials.

50. While not part of the 2019 Amgen Retrospective Study, Amgen conducted yet another retrospective study using the MarketScan data, entitled "Estimating the Association of Patient Characteristics and Environmental Factors with G-CSF Use and Outcomes" ("Retrospective Factors Study"), that shows that compliance is an especially important consideration with regards to treatment in this context.[38] The Retrospective Factors Study concludes that cycles corresponding to patients with metastatic cancer, longer drive times, or cycles falling on a day before the weekend or a holiday have statistically lower compliance rates.[39]

51. In the 2019 Amgen Retrospective Study, Amgen also performed additional sensitivity analyses to compare the OBI and PFS incidence rate of FN by cycle and by FN risk group.[40] These sensitivity analyses demonstrate results consistent with those obtained in the primary study analysis, thereby strengthening the primary analysis finding of a statistically significant association between OBI and reduced FN incidence rate. Cycles 1 through 5 show higher risk of FN in patients using PFS relative to OBI (RRs between 0.54 and 0.88) with cycle 4 showing a statistically significant increase in risk of FN among PFS uses (RR = 0.54 and p = 0.03).[41] Amgen also conducted these analyses for high-risk cycles, again finding results consistent with those obtained in the primary study analysis.[42] In addition, the 2019 Amgen Retrospective Study included an analysis at the patient rather than cycle level. This sensitivity analysis again produced consistent results (RR = 0.85 and p = 0.16).[43]

52. Finally, the 2019 Amgen Retrospective Study contains two additional sets of analyses: one for patients without secondary malignancy and one for patients with secondary malignancy.[44] The results were again consistent with those corresponding to the primary analyses. For cycles among patients without secondary malignancy, the OBI and PFS FN

[38] The study has very similar inclusion criteria. See Deposition of Mark Hatfield, April 13, 2023 ("Hatfield Deposition"), Exhibit 53, at AMG10939-0000004074.
[39] Hatfield Deposition, Exhibit 53, at AMG10939-0000004094.
[40] The risk level was defined based on the regimen. Among intermediate risk cycles, additional risk factors were identified both at the cycle- and patient-level. See AMG10939-0000303957–84 at 67.
[41] AMG10939-0000304081.XLSX, tab "5a_FN_Cycle". See also Workpaper 1, Rows 002–006, and my backup materials. For cycle 6, the results are reversed, but the estimate of relative risk is not statistically different from 1 (RR = 1.45; p = 0.33). See Workpaper 1, Row 007, and my backup materials. The primary reason why the individual cycle analyses lead to results that are not statistically significant is that the sample size for each individual cycle is much smaller than the sample size for the data pooled across all cycles. The smaller sample sizes result in underpowered estimates, which I explain in more detail below when I present and discuss the results of the 2021 Sandoz Retrospective Study.
[42] AMG10939-0000303957–84 at 73. For the high-risk cycles only RR = 0.79 and p = 0.05. See AMG10939-0000304080, tab "6_FN@CYC_Hi Risk". See also Workpaper 1, Row 008, and my backup materials.
[43] AMG10939-0000304080, tab "5b_FN&Cost_Patient". See also Workpaper 1, Row 009, and my backup materials.
[44] AMG10939-0000303957–84 at 73.

B. The 2021 Sandoz Retrospective Study

54. Sandoz conducted a study similar to the 2019 Amgen Retrospective Study using MarketScan data to compare OBI and PFS in terms of FN rates.[50] The 2021 Sandoz Retrospective Study was funded by Sandoz, and the authors were employees, contractors, or consultants to Sandoz.[51] The initial dataset consisted of 3,152 patients and 11,196 cycles.[52] Using 1:1 propensity score matching,[53] the authors matched 2,170 of the 3,152 patients (1,085 OBI and 1,085 PFS) with 7,467 cycles.[54] Both 1st cycle and all-cycle analyses were conducted.[55]

55. The authors found that for the 1st cycle analysis the adjusted (matched sample) rates of FN were 1.01% for OBI and 1.48% for PFS .[56] The authors report that this difference was not statistically significant, and my re-analysis of these data confirm a non-significant 45% increase in the rate of FN in PFS treated patients relative to OBI treated patients (RR = 0.69; CI = 0.32, 1.47; p = 0.34).[57] For all-cycle analyses, the rate of FN for OBI was 0.91% versus 1.22%, and again, the authors report that this difference was not statistically significant.[58] Because the authors do not report the sample sizes for the all-cycles analysis, I cannot compute the 95% confidence interval for the RR using the same technique as I use throughout the rest of this report. The observed RR is 0.91/1.22 = 0.75, or a 33% increase in rate.

56. If the same all-cycle RR was found for the sample size (number of cycles) in the 2019 Amgen Retrospective Study, the adjusted (matched) analysis would have found this 33% increase in FN rate to be statistically significant (RR = 0.75, CI = 0.59, 0.95; p=0.01).[59] It is not the adjustment that yielded a non-statistically significant result, but the reduced sample size of the 2021 Sandoz Retrospective Study that led to their failure to detect a statistically

---

[50] 2021 Sandoz Retrospective Study, at p.1230.
[51] 2021 Sandoz Retrospective Study, p. 1237.
[52] 2021 Sandoz Retrospective Study, p. 1234.
[53] 1:1 propensity score matching attempts to balance the sample on observable risk by finding for each OBI patient a PFS patient that has a similar probability of developing FN after controlling for observable patient characteristics. Because the original number of PFS patients is almost double the number of OBI patients, such a 1:1 design inevitably discards approximately one third of all observations.
[54] 2021 Sandoz Retrospective Study, p.1234.
[55] 2021 Sandoz Retrospective Study, p.1234.
[56] 2021 Sandoz Retrospective Study, p.1234.
[57] 2021 Sandoz Retrospective Study, p.1234. See also Workpaper 1, Row 012, and my backup materials.
[58] 2021 Sandoz Retrospective Study, p.1234.
[59] In conducting this analysis, I used the number of cycles for the OBI and PFS cohorts in the 2019 Amgen Retrospective Study and applied the corresponding rates of FN incidence for all-cycle analyses observed in the 2021 Sandoz Retrospective Study to those two groups. See Workpaper 1, Row 013, and my backup materials.

significant result. This is further confirmed by their finding of a smaller difference for the unadjusted analysis (0.81% for OBI versus 1.00% for PFS, RR = 0.81/1.00 = 0.81).[60] Without adjustment (matching) the difference is a 23% increase in rate of FN for PFS, whereas with adjustment there is a 33% increase in rate of FN. The Sandoz data reveal that statistical adjustment based on the propensity score actually increases the magnitude of the benefit derived from OBI.

57. Based on the results of the 2021 Sandoz Retrospective Study, the authors conclude that "[i]n a matched cohort of patients representing real-world utilization, there was no statistically or clinically meaningful difference in FN incidence between OBI and PFS methods of pegfilgrastim."[61] This is a misleading and inappropriate statement given the results of the study. The study was not powered to detect a clinically meaningful difference, and the estimated difference of a 45% increase in 1st cycle FN events and a 33% increase in all-cycle FN events are far from trivial.[62] I have designed numerous RCTs and observational studies that are powered to detect relative differences of 25% in incidence rates of different outcomes. Sandoz could have used an alternative propensity score matching algorithm, which would not have tossed out observations from the analysis, and thus would have resulted in a study that was more adequately powered. In fact, the 2021 Amgen Prospective Study used an inverse propensity score weighting, which is the optimal way of implementing propensity score matching without throwing away any observations. By retaining the full sample, the 2021 Amgen Prospective Study had sufficient statistical power to detect a statistically significant effect for a similar reduction in the FN incidence rates.

58. To shed further light on the sample size issue, I perform a series of calculations that quantify how underpowered the 2021 Sandoz Retrospective Study actually was. First, using the study's actual sample size, I compute the power of that study in detecting the estimated difference in the FN incidence rates. For the 1st cycle analysis, which analyzed a sample of 2,170 patients, the power is only 11.64%, and if the FN rates from the all-cycles analyses were analyzed instead, the power is even lower—6.26%.[63] Next, I calculate the sample size that would be required to guarantee adequate power assuming the rest of the study design is unchanged. Using the FN incidence rate estimates from the 1st cycle and all-cycles analyses,

---

[60] 2021 Sandoz Retrospective Study, p.1234.
[61] 2021 Sandoz Retrospective Study, p.1230.
[62] In fact, Amgen considered a relative difference of 25% in the FN incidence rates as the "minimally important difference" when designing its prospective study. See AMG10939-0000301501–15 at 05.
[63] See my backup materials. Because the authors do not report the sample sizes for the all-cycles analysis, I cannot compute power using the same methodology.

(2013), this is the "gold standard" definition of FN within the literature analyzing this illness.[69]

61. The primary analysis included N = 2,575 (OBI = 1,624, Other = 951) subjects out of 2,715 subjects enrolled in the study.[70] The comparator group, referred to as the "Other group" in the study included: (i) subjects that received G-CSF (granulocyte-colony stimulating factor) prophylaxis (69.1%), through the administration of Neulasta pre-filled syringe (PFS), pegfilgrastim biosimilar PFS, filgrastim, tbofilgrastim, "filgrastim-sndz," and (ii) subjects who received no G-CSF prophylaxis (30.9%) selected at the physician's discretion in the first cycle.[71] According to the study protocol, the patient's group was defined by the prophylaxis option administered in the first cycle and remained constant throughout the whole course of the study even if other prophylaxis options were administered in subsequent cycles.[72] The study performs secondary analyses which compare the OBI cohort to the Neulasta PFS or biosimilar PFS subset of the Other group.[73]

62. In general, the groups were similar at baseline with the exception of surgery within the past 6 months, for which the rate was higher in the OBI group (77%) compared to the Other group (63%).[74] With the exception of gender, the distribution of baseline patient demographics and characteristics remained consistent across groups both pre- and post-COVID.[75] Despite a decrease in sample size, the overall share of female subjects increased post-COVID (79.1% pre-COVID to 89.3% post-COVID).[76] The results for secondary analyses adjusting for covariates were reported in the study and demonstrated similar results. Covariate adjustment was based on inverse proportional weighting (standardized marginal structural log-binomial model).[77] This is similar to propensity score matching. However,

[69] Derek Weycker et al., "Technical Evaluation of Methods for Identifying Chemotherapy-Induced Febrile Neutropenia in Healthcare Claims Databases," *BMC Health Services Research*, 13(60), 2013, pp. 1–10 ("Weycker, et al. (2013)"), at p. 3 ("The gold standard for identification of FN hospitalization (i.e., presumptive, for purposes of this retrospective evaluation) was evidence in EMR data of ANC <1.0 × 10^9/L and either body temperature ≥38.3°C (101°F) or administration of antibiotic or antiviral therapy following ANC assessment, all having occurred within 1 day of hospitalization (i.e., the day before, day of, or day after hospitalization); this definition was considered the presumptive "gold standard" for purposes of analyses.").
[70] AMG10939-0000302246–3296 at 2261.
[71] AMG10939-0000302246–3296 at 2351.
[72] AMG10939-0000302246–3296 at 2278.
[73] AMG10939-0000302246–3296 at 2358. However, most of the results presented in the 2021 Amgen Prospective Study are not broken down at the level of any specific sub-group within the Other group.
[74] See AMG10939-0000302246–3296 at 2317–2319 for the fuller discussion of study baseline demographic characteristics and AMG10939-0000302246–3296 at 2320–2325 for the study baseline characteristics.
[75] AMG10939-0000302246–3296 at 2414–2417 and 2504–2515, Tables 14-2.1.7, 14-2.1.8, 14-2.2.403, and 14-2.2.404.
[76] AMG10939-0000302246–3296 at 2414–2416, Tables 14-2.1.7 and 14-2.1.8.
[77] AMG10939-0000302246–3296 at 3108.

unlike the 1:1 matching algorithm used in the 2021 Sandoz Retrospective Study that required that the investigators discard roughly a third of the sample due to inadequate available matches, this weighting approach allows inferences to be based on the full sample (no discards), preserving the representativeness of the sample.[78]

63. The results of the 2021 Amgen Prospective Study revealed that OBI administration in the 1st cycle was associated with a decreased FN rate across all cycles. FN rates were 6.4% (CI = 5.21%, 7.59%) for OBI and 9.4% (CI = 7.51%, 11.21%) for the Other group, corresponding to a rate increase of 47% (RR = 0.68; CI = 0.52, 0.90; p = 0.007).[79] Restricting the analysis to patients who received OBI in every cycle further decreased the rate in the OBI group, leading to a 52% increase in FN for the Other group (RR = 0.66; CI = 0.50, 0.88; p = 0.004).[80] After controlling for potential confounders, the study results demonstrate a further reduction in FN rates for the OBI group. Adjusted rates of FN were 6.6% (CI = 5.28%, 8.04%) for OBI and 9.9% (CI = 7.52%, 12.33%) for the Other group, corresponding to a 52% increase in the incidence of FN (RR = 0.66; CI = 0.47, 0.91; p = 0.006).[81] For subjects who received OBI in every cycle, the adjusted rate of FN was 6.3% (CI = 4.9%, 7.6%), corresponding to a 56% increase in the incidence of FN relative to the Other group (RR = 0.64; CI = 0.46, 0.85; p = 0.004).[82] Compliance (administration the day after chemotherapy ended) was also significantly higher in the OBI group (88.3%, CI = 86.7%, 89.9%) versus the Neulasta PFS or biosimilar PFS subset of the Other group (48.8%, CI = 45.0%, 52.6%), with non-overlapping confidence intervals denoting statistical significance.[83]

64. Furthermore, for patients who received pegfilgrastim the day after chemotherapy ended ("compliers"), FN incidence was similar between the OBI and the Neulasta PFS or biosimilar PFS groups. For compliant subjects, FN incidence was small and similar in both groups, with OBI having a 6.3% (CI = 5.0%, 7.5%) FN rate relative to 6.8% (CI = 4.1%, 9.6%) for the Neulasta PFS or biosimilar PFS subset of the Other group with corresponding

---

[78] For example, the matched cohort had a notably narrower age range (46-59) as compared to the full sample (23-88) and was generally younger (4.8% vs. 9.6% of patients 65 years or older). See 2021 Sandoz Retrospective Study, p. 1233.
[79] AMG10939-0000302246–3296 at 2329–2330, Table 8. See Workpaper 1, Row 014, and my backup materials.
[80] AMG10939-0000302246–3296 at 2329–2330, Table 8. See Workpaper 1, Row 015, and my backup materials.
[81] AMG10939-0000302246–3296 at 2588, Table 14-4.2.22.
[82] AMG10939-0000302246–3296 at 2595, Table 14-4.3.400.
[83] Compliance is defined as receipt of Neulasta Onpro, Neulasta PFS, or pegfilgrastim biosimilar PFS the day after completion of chemotherapy, in all cycles in which pegfilgrastim is administered. Note that the compliance analysis uses the Neulasta PFS and PFS biosimilars subset of the Other group as the comparator group. AMG10939-0000302246–3296 at 2566, Table 14-4.1.402, Table 17.

RR=0.92 (CI = 0.59, 1.45; p = 0.71).[84] The incidence of FN was higher for non-compliers than for compliers in both groups, with the rate of FN of 7.4% (CI = 3.7%, 11.1%) in the OBI group and 12.1% (CI = 8.6%, 15.6%) in the Neulasta PFS or biosimilar PFS subset of the Other group, resulting in RR=0.61 (CI = 0.34, 1.09; p = 0.09).[85]

65. The 2021 Amgen Prospective Study also disaggregates FN incidence rates by sub-groups that received G-CSF prophylaxis in all, at least one, and no cycles within the Other group. For subjects that received G-CSF in all cycles, FN incidence was 7.8% (CI = 5.8%, 9.9%).[86] The rate of FN increases for subjects that received G-CSF support for at least one, but maybe not all, cycles to 10.0% (CI = 7.8%, 12.1%).[87] Interestingly, the rate of FN reduced to 7.5% (CI = 4.1%, 10.9%) among subjects that did not receive G-CSF support in any cycle.[88] In the real world, patients may not be compliant with their treatment program. For patients who received G-CSF prophylaxis in some *but not all* cycles, not receiving G-CSF prophylaxis in some cycles could be related to compliance rather than a lack of need of treatment. Thus, the inclusion of these patients in the Other sample provides relevant insights with respect to the real-world performance of different prophylaxis methods. Furthermore, the lower rate of FN in control subjects that did not receive G-CSF support suggests that they were at lower risk of developing FN to begin with, on the theory that their physicians often adopt a "watch and wait" secondary prophylaxis approach.[89] Their inclusion in the Other group therefore provides a downward bias in the difference in FN incidence between the OBI and the Other group, such that the observed difference provides a lower-bound on the true, potentially greater difference in favor of OBI.

66. The 2021 Amgen Prospective Study further distinguishes between different cancer types in conducting its analysis.[90] For each cancer type studied, the groups under comparison

[84] AMG10939-0000302246–3296 at 2566, Table 14-4.1.402. See also Workpaper 1 Row 016, and my backup materials.
[85] AMG10939-0000302246–3296 at 2568, Table 14-4.1.403. See Workpaper 1, Row 017, and my backup materials.
[86] AMG10939-0000302246–3296 at 2562, Table 14-4.1.400b.
[87] AMG10939-0000302246–3296 at 2561, Table 14-4.1.400a.
[88] AMG10939-0000302246–3296 at 2563, Table 14-4.1.400c.
[89] "Watch and Wait," *Leukemia & Lymphoma Society*, available at https://www.lls.org/treatment/types-treatment/watch-and-wait, accessed on May 8, 2023 ("Watch and wait involves closely monitoring a patient's condition without giving any treatment until symptoms appear or change.").
[90] The 2021 Amgen Prospective Study presents a breakdown of certain analyses by the four cancer types: breast cancer, lung cancer, prostate cancer, and NHL.

continues."[106] I have reviewed whether the claims in these marketing materials are consistent with the findings in the 2021 Amgen Prospective Study. The marketing materials correctly summarize the study design and findings by stating: "[i]n a prospective observational study of ~2600 cancer patients[, f]ewer patients experienced FN with Neulasta® Onpro® vs other FN-prophylaxis options." These materials also accurately represent incidence rates of FN in the different study arms, corresponding confidence intervals, sample sizes, and the results of the adjusted study analyses.[107] The marketing materials also disclose limitations of the study. For example, the marketing materials indicate "it was not possible to evaluate FN risk among patients lost to follow-up after study enrollment."[108] Also, since this study was not an RCT, the marketing materials clearly state that the "lack of randomization means that the groups may have differed in ways that were not measured or recorded. The impact of such differences on the study findings is unknown."[109] I conclude that the 2021 Amgen Prospective Study provided valuable statistical insights of scientific interest, and as such, it was reasonable for Amgen to rely on the 2021 Amgen Prospective Study in issuing these marketing materials.

D. Meta-analysis of the Amgen and Sandoz Studies Provides Further Statistical Evidence of a Lower Rate of FN Incidence Associated with the Use of OBI

71. The two Amgen studies find a statistically significant association between OBI and rate of FN relative to PFS and other comparators, and the 2019 Sandoz Retrospective study found a similar effect that was *not* statistically significant. The vote count is two to one in favor of statistically significant results.[110] However, a better approach than vote counting (when there is a consistent outcome, e.g., FN) is to synthesize the results across the available studies using meta-analysis. Meta-analysis pools summary statistics (e.g., relative risk point

[106] Complaint, ¶ 92.
[107] Complaint, ¶ 96, Exhibit 6.
[108] Complaint, Exhibit 6.
[109] Complaint, Exhibit 6.
[110] According to the Cochrane Collaboration, "vote counting might be considered as a last resort in situations when standard meta-analytical methods cannot be applied (such as when there is no consistent outcome measure)." See Jonathan J, Deeks, Julian P.T. Higgins, and Douglas G Altman, "9.4.11 Use of vote counting for meta-analysis" of *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0, ed. Julian P.T. Higgins and Sally Green: The Cochrane Collaboration, 2011, available at https://handbook-5-1.cochrane.org/chapter_9/9_4_11_use_of_vote_counting_for_meta_analysis.htm. The Cochrane Collaboration is an international, not-for-profit organization that aims to help people make well-informed decisions about health care by preparing, maintaining, and promoting the accessibility of systematic reviews of the effects of health-care interventions. See "About Us," *Cochrane*, available at https://www.cochrane.org/about-us.

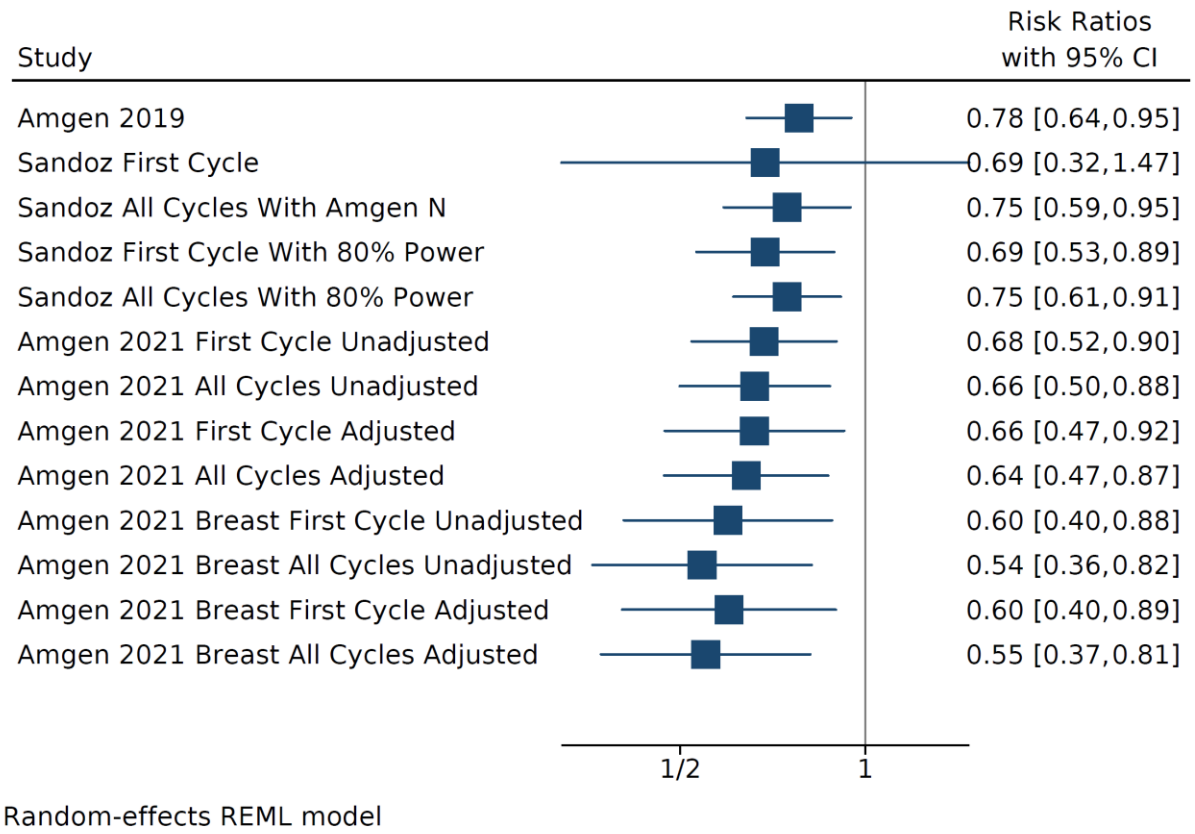
estimates and their 95% confidence intervals) across the available studies to come up with a summary of the overall effect (e.g., a pooled relative risk, inversely weighted by uncertainty in the individual studies), uncertainty in the pooled estimate (i.e., a 95% confidence interval), and a test of the overall null hypothesis of no association (a p-value). Random-effect meta-analysis permits variability in the treatment effect over studies and fixed-effect meta-analysis assumes that the effect of treatment is constant over the studies. Statistical tests are performed to determine if the treatment effect varies over studies. However, the safest approach is to assume that it does and use a random-effect model.

72. A major advantage of meta-analysis is that it benefits from the increased sample size derived from pooling multiple studies. For example, it is possible that none of a series of small studies found a statistically significant treatment effect, but the relative risks were all approximately 0.5, indicating a 100% increase in risk associated with PFS. A meta-analysis could well find a statistically significant treatment effect by benefiting from a consistent treatment effect across all of the studies and a substantially larger sample size. Meta-analyses are widely used in consensus building in medical research.[111]

73. In order to synthesize the information across the three studies, I have performed a random-effects meta-analysis looking at FN as the outcome. In Exhibit 1, I provide a forest plot of the relative risk estimates and their 95% confidence intervals for the 2019 Amgen Retrospective Study, the 2021 Sandoz Retrospective Study, and the 2021 Amgen Prospective Study as well as some sensitivity analyses discussed in my report. The RRs range from 0.54 to 0.78, and all but the 2021 Sandoz Retrospective Study are statistically significant (i.e., 95% confidence intervals do not contain the value of 1.0).

[111] As noted by Cochrane Collaboration "If the results of the individual studies are combined to produce an overall statistic, this is usually called a meta-analysis. Many Cochrane Reviews measure benefits and harms by collecting data from more than one trial, and combining them to generate an average result. This aims to provide a more precise estimate of the effects of an intervention and to reduce uncertainty." See "About Cochrane Reviews," *Cochrane*, available at https://www.cochranelibrary.com/about/about-cochrane-reviews.

EXHIBIT 1 RR for All Studies



Source: 2019 Amgen Retrospective Study; 2021 Sandoz Retrospective Study; 2021 Amgen Prospective Study; backup materials

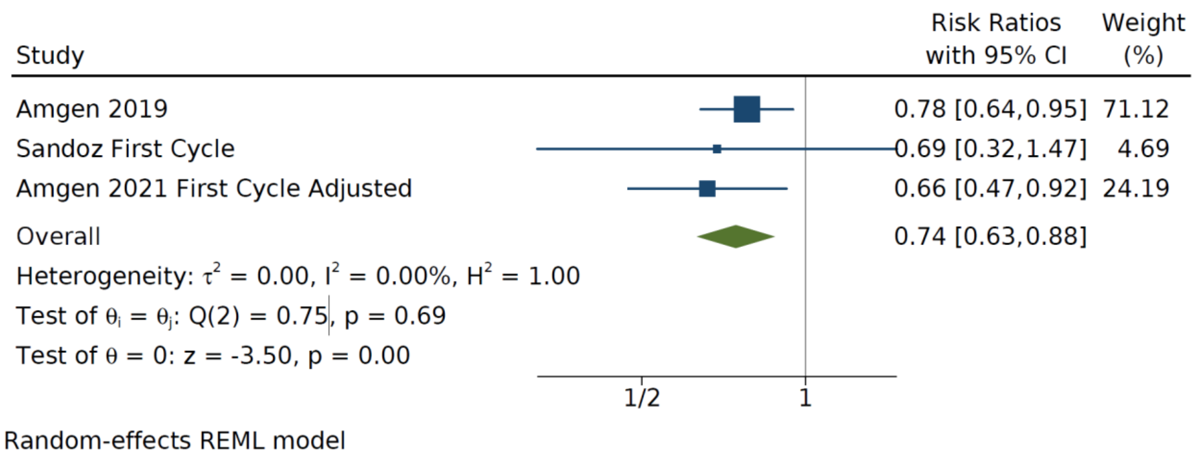
Note: The blue squares represent the RR for each analysis. The horizontal blue lines depict the corresponding CIs. The "Sandoz First Cycle with 80% Power" and "Sandoz All Cycles with 80% Power" calculations present the RRs and CIs based on the sample sizes required to achieve 80% power which I calculated previously.

74. In Exhibit 2, I display the results of a random-effect meta-analysis pooling the RRs across the three studies (the 2019 Amgen Retrospective Study, the 2021 Sandoz Retrospective Study, and the 2021 Amgen Prospective Study with all cancer types, OBI in the 1st cycle, and adjusted FN incidence rates). The pooled estimate of the relative risk is RR = 0.74 (CI = 0.63, 0.88; p = 0.0005). This estimate reveals that when pooling the data from all three studies, including Sandoz's own study, the overall result (incorporating variability in the treatment effect) is <u>a statistically significant 35% increase in the rate of FN</u>.[112] This is a

[112] Despite differences in design protocols across the three studies, the random-effects model finds no indication of a heterogenous treatment effect, as seen by the three studies' similar point estimates and $\tau^2 = 0.00$. To account for potential correlation in FN events across repeated cycles in the 2019 Amgen Retrospective Study, I perform a sensitivity analysis in which I exclude the 2019 Amgen Retrospective Study. The pooled estimate of the relative risk across the Sandoz and 2021 Amgen Prospective Studies is RR = 0.66 (CI = 0.49, 0.90; p = 0.01). This estimate reveals that pooling the data across two of the three studies results in a slightly wider confidence interval, but still finds a large and statistically significant FN incidence rate increase of 52%. Notably, this increase is even greater than that obtained from pooling the three studies (35%). See Workpaper 2.

larger effect than the effect found in the 2019 Amgen Retrospective Study and reported in the corresponding set of marketing materials (31%).

EXHIBIT 2RR for Three Unique Studies



Source: 2019 Amgen Retrospective Study; 2021 Sandoz Retrospective Study; 2021 Amgen Prospective Study; backup materials

Note: The blue squares represent the RR for each analysis. The sizes of the squares are proportional to the weight of a given study in the meta-analysis. The horizontal blue lines depict the corresponding CIs.

VII. The Criticisms of the Amgen Studies Raised by Sandoz in the Complaint Fail to Undermine the Statistical Evidence of a Lower Incidence of FN Associated with the Use of OBI

75. In this section, I review the different criticisms of the Amgen studies made by Sandoz in the Complaint. I explain how each of these criticisms is either immaterial in this setting or how they have already been addressed by various robustness measures undertaken in the Amgen and Sandoz studies. None of these criticisms undermine the statistical evidence of lower rates of FN associated with OBI.

76. *Amgen's claims regarding the lower rate of FN with OBI relative to PFS were not based on a clinical trial, but rather a retrospective study* (Complaint, ¶ 57). Amgen did not claim that the 2019 Amgen Retrospective Study was a clinical trial. As discussed earlier, observational studies can provide valuable statistical evidence with respect to the evaluation of a given treatment. The 2019 Amgen Retrospective Study provides evidence of an association between OBI and reduced rates of FN.

77. *The Complaint states that Amgen did not perform a comparison with a PFS biosimilar in the 2019 Amgen Retrospective Study* (Complaint, ¶ 58). For the 2019 Amgen Retrospective Study, biosimilars were not included because they were not found in the

- Although not statistically significant due to a small sample size (partially caused by the use of a statistical matching algorithm that required excluding approximately one-third of the data), the 2021 Sandoz Retrospective Study found effect sizes even larger than the 2019 Amgen Retrospective Study. For the 1st cycle analysis, there was a 45% increase in FN rate for patients treated with PFS versus OBI (RR = 0.69; CI = 0.32, 1.48; p = 0.34). If the same sample size that was used in the 2019 Amgen Retrospective Study were used in the 2021 Sandoz Retrospective Study, the result would have been a statistically significant 33% increase in FN incidence rate associated with the use of OBI relative to PFS (RR = 0.75; CI = 0.59, 0.95; p = 0.01).
- Finally, the 2021 Amgen Prospective Study, which used a prospective design with the gold standard biological/clinical definition of FN and adjusted for potential confounders through inverse proportional treatment weighting in its secondary analyses, identified an even larger adjusted reduction in the FN incidence rate associated with the use of OBI versus other prophylaxis options in 1st cycle treatment (RR = 0.66; CI = 0.47, 0.92; p = 0.003) and across all cycles (RR = 0.64; CI = 0.46, 0.85; p = 0.004). Importantly, both the 2021 Sandoz Retrospective Study and the 2021 Amgen Prospective Study found *even larger reductions* in the FN incidence rates than the 2019 Amgen Retrospective Study, after controlling for observed confounders. This provides further statistical evidence in support of a reduction in FN incidence rate associated with the use of OBI.

96. Synthesis of the results from all three studies in a meta-analysis described above reveals a statistically significant 35% increase in FN rate in patients treated with other prophylaxis methods (including PFS, PFS biosimilars, and no G-CSF prophylaxis) versus OBI (RR = 0.74; CI = 0.63, 0.88; p=0.0005). Based on the results of the three studies discussed in this report, this is the best estimate of the magnitude of the FN rate reduction effect of OBI relative to PFS as it pools data from all three studies. As such, the marketing claims made by Amgen that highlight the reduction in FN rate associated with the use of OBI are therefore consistent with the findings of both Amgen studies. Thus, the Amgen studies provide valuable statistical insights of scientific interest, such that it was reasonable for Amgen to issue these marketing materials based on the results of these studies.